ICS 35.020
CCS L 70

NATIONAL STANDARD OF THE PEOPLE'S REPUBLIC OF CHINA

GB/T 45280-2025

Artificial intelligence - Unified interfaces for heterogeneous artificial intelligence accelerating units

人工智能 异构人工智能加速器统一接口

Issued on: February 28, 2025 **Implemented on: February 28, 2025**

Issued by: State Administration for Market Regulation;
Standardization Administration of the PRC

Table of Contents

Foreword

This document is in accordance with the provisions of GB/T 1.1-2020 "Guidelines for standardization work Part 1: Structure and drafting rules for standardization documents" Drafting.

Please note that some of the contents of this document may involve patents. The issuing organization of this document does not assume the responsibility for identifying patents.

This document was proposed and coordinated by the National Technical Committee for Information Technology Standardization (SAC/TC28).

This document was drafted by: China Electronics Technology Standardization Institute, Huawei Technologies Co., Ltd., Shanghai Artificial Intelligence Innovation Center, Beijing Aerospace University of Aeronautics and Astronautics, Shanghai Enflame Technology Co., Ltd., Beijing BiRen Technology Development Co., Ltd., Shanghai Tianshu Zhixin Semiconductor Co., Ltd., Cambrian Technologies Co., Ltd., Intel (China) Co., Ltd., Shenzhen Intellifusion Technology Co., Ltd., Muxi Integrated Circuit Co., Ltd.

Road (Shanghai) Co., Ltd., Institute of Software, Chinese Academy of Sciences, Inspur Electronic Information Industry Co., Ltd., Peking University, Shanghai Artificial Intelligence Energy Industry Association, Huawei Cloud Computing Technology Co., Ltd., Shanghai SenseTime Intelligent Technology Co., Ltd., Beijing Zhixin Microelectronics Technology Co., Ltd., Pingtouge (Shanghai) Semiconductor Technology Co., Ltd., Hangzhou Hikvision Digital Technology Co., Ltd., China Mobile Communications Group Co., Ltd.

Shenzhen Kunyun Information Technology Co., Ltd., Loongson Technology Co., Ltd., Nanjing NARI Ruiteng Technology Co., Ltd., Suzhou Deng Lin Technology Co., Ltd., China Southern Power Grid Artificial Intelligence Technology Co., Ltd., Midea Group (Shanghai) Co., Ltd., Peking University Changsha Computing and Digital Economy Research Institute, Beijing University of Aeronautics and Astronautics Hangzhou Innovation Research Institute, Sichuan Huakun Zhenyu Intelligent Technology Co., Ltd., Shandong Inspur Technology China Southern Power Grid Co., Ltd., Southwest University of Science and Technology, Zhejiang Dahua Technology Co., Ltd., Beijing Greenland Shenzhen Tong Information Technology Co., Ltd. and Beijing Electronic Digital Technology Co., Ltd.

The main drafters of this document are: Dong Jian, Yang Yuze, Zhang Yali, Xu Yang, Zhang Chengxing, Bao Wei, Liu Wenfeng, Pei Zhilin, Wang Waner, Cao Xiaoqi, Ma Chenghao, Luan Zhongzhi, Mei Jingqing, Ding Ruiquan, Hu Mingshan, Cheng Guipeng, Wang Haining, Su Lan, Liu Zi, Meng Lingzhong, Ma Shanshan, Li Binbin, Su Dongdong, Yang Chao, Zhao Chunhao, Liu Yong, Zhong Pu, Zhang Yibo, Zhang Fang, Jin Di, Cai Quanxiong, Ma

Wanyue, Shi Chao, Ci Hongbin, Chen Rouyi, Cai Yasen, Jia Mengzhu, Hu Zheng-hui, Zhao Yan-jun, Li Rui, Zhang Xi-ming, Xu Huan, Yu Wen-xin, Fang Gui-ming, Yu Jie, Guo Wen.

Artificial Intelligence Heterogeneous Artificial Intelligence Accelerator Unified Interface

1 Scope

This document defines the unified interface of heterogeneous artificial intelligence accelerators, its semantics and usage, and describes the functions of each accelerator to implement this interface.

The access methods and test methods required for the port.

This document is applicable to the design and implementation of artificial intelligence accelerator interfaces, and can also provide a reference for artificial intelligence accelerator applications.

2 Normative references

GB/T 41867-2022

GB/T 45087-2024

GB/T 17966-2024

3 Terms and definitions

The terms and definitions defined in GB/T 41867-2022 and the following apply to this document.

3.1 [application programming] interface

< artificial intelligence> Syntax and semantic definitions for using artificial intelligence accelerator functions.

[Source. ISO /IEC /IEEE9945 2009, 3.19, modified]

3.2 artificial intelligence accelerating processor

artificial intelligence accelerating chipartificialinteligenceacceleratingchip An integrated circuit component with a computing microarchitecture that is adapted to artificial intelligence algorithms and can complete the computing and processing of artificial intelligence applications.

Note. In this document, AI accelerator is referred to as accelerator in a context that does not cause misunderstanding.

[Source. GB/T 41867-2022, 3.1.5]

3.3 computationalgraph

A directed graph consisting of nodes and links used to represent mathematical functions.

Note 1.A node represents a mathematical operation, i.e. an operator.

Note 2.Connections represent dependencies between mathematical operations.

Note 3.A connection connects the start node and the end node.

[Source. ISO /IEC /IEEE24765 2017, 3 1762.1, modified]